Data Processing Subsystem

Mark Shannon

mshannon@eos.hitc.com

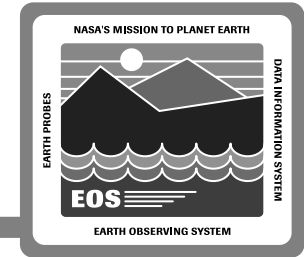**ECS Release A SDPS/CSMS Critical Design Review**
16 August 1995

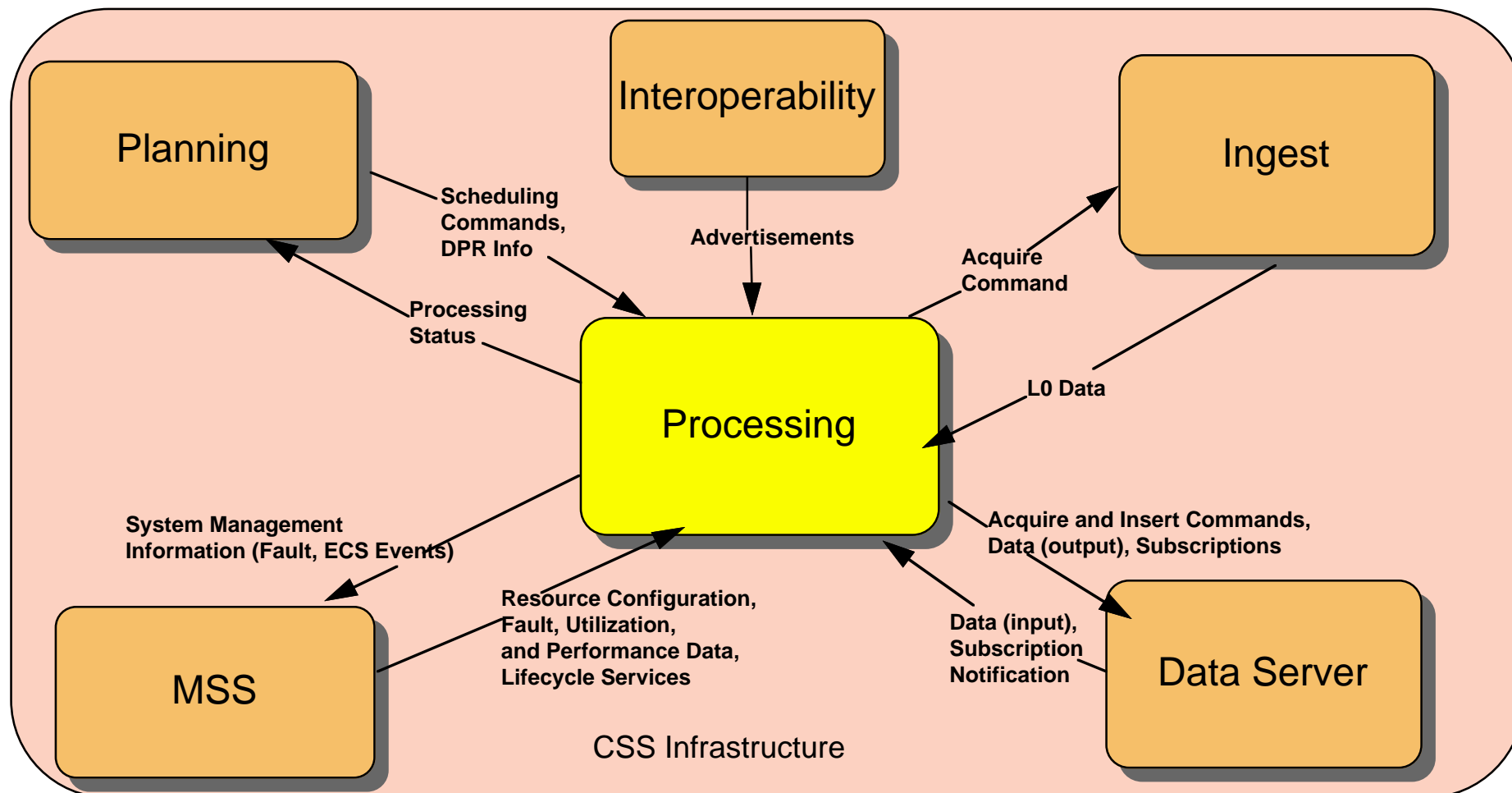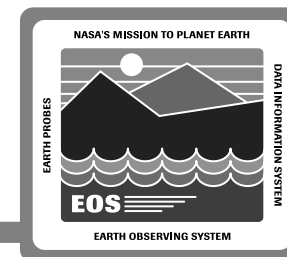# Processing CSCI Topics

- **Overview**
- **Context**
- **CSCs and Relationships**
- **Data Processing Subsystem Hardware Architecture**
- **Software to Hardware Mapping**
- **Algorithm Quality Assurance HW**
- **Production Scenario**
- **Evolvability**
- **Issues and Recommendations**

Processing CSCI Overview

- **Manage and monitor the production of Science Data Products in a distributed, heterogeneous UNIX environment.**

- **Provide Operations Interfaces to allow intervention into automated production, when required.**

- **Provide interface to the Data Server to support staging and destaging of data.**

- **Provide fault and fault recovery capabilities to support science data production.**

- **Manage the use of Science Processing Hardware Resources to support Science Data production.**

Processing CSCI Interfaces

NASA'S MISSION TO PLANET EARTH

EARTH PROBES

DATA INFORMATION SYSTEM

EOS

EARTH OBSERVING SYSTEM

Planning

Interoperability

Ingest

Scheduling Commands, DPR Info

Advertisements

Acquire Command

Processing Status

Processing

L0 Data

System Management Information (Fault, ECS Events)

Acquire and Insert Commands, Data (output), Subscriptions

MSS

Resource Configuration, Fault, Utilization, and Performance Data, Lifecycle Services

Data (input), Subscription Notification
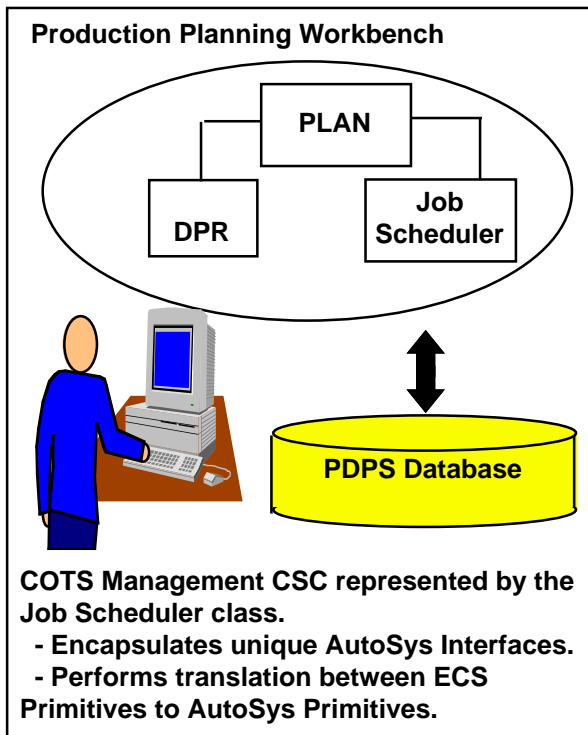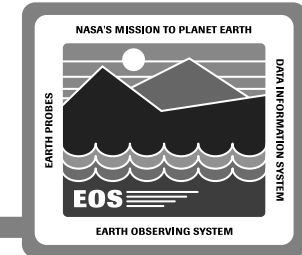
Data Server

CSS Infrastructure

Processing CSCs

- COTS
- COTS Management
- Data Management
- PGE Execution Management
- Resource Management
- Data Pre-Processing
- QA Monitor

A Processing CSC is mapped to one of the following types :
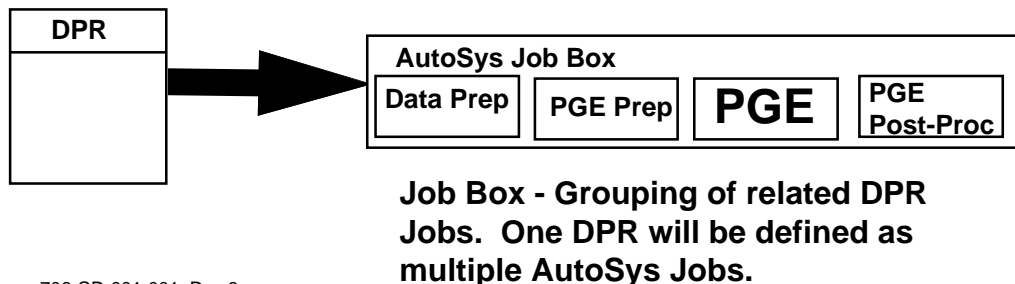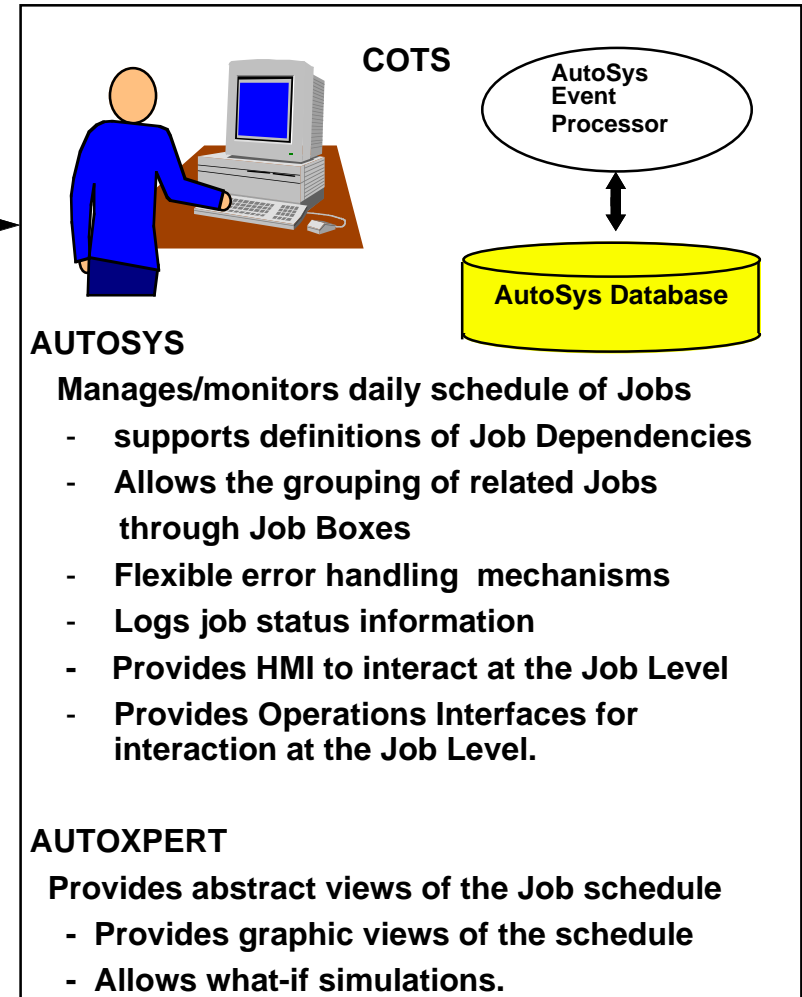1) Application - executable used to perform an ECS specific function
2) Class Library - a low-level collection of classes used by different applications.

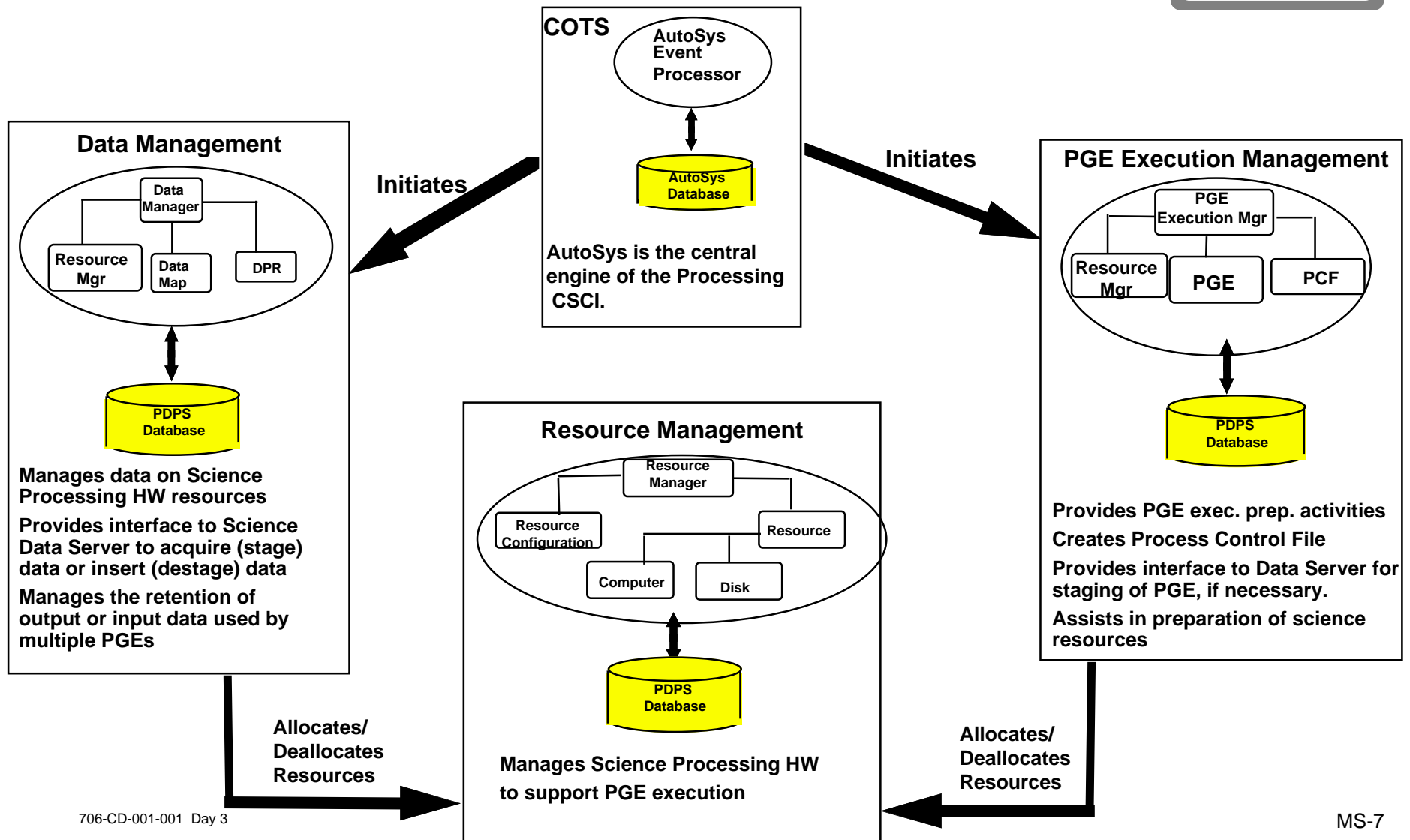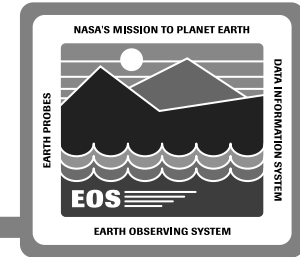Processing persistent data is managed through the PDPS Database.
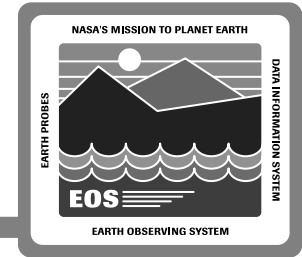
Planning and Processing CSC Interfaces

Production Planning Workbench

PLAN

DPR

Job Scheduler

PDPS Database

COTS Management CSC represented by the Job Scheduler class.
- Encapsulates unique AutoSys Interfaces.
- Performs translation between ECS Primitives to AutoSys Primitives.

Job Information

COTS

AutoSys Event Processor

AutoSys Database

AUTOSYS

Manages/monitors daily schedule of Jobs

- supports definitions of Job Dependencies
- Allows the grouping of related Jobs through Job Boxes
- Flexible error handling mechanisms
- Logs job status information
- Provides HMI to interact at the Job Level
- Provides Operations Interfaces for interaction at the Job Level.

AUTOXPERT

Provides abstract views of the Job schedule

- Provides graphic views of the schedule
- Allows what-if simulations.

DPR

AutoSys Job Box

| Data Prep | PGE Prep | PGE | PGE Post-Proc |

Job Box - Grouping of related DPR Jobs. One DPR will be defined as multiple AutoSys Jobs.

Processing CSCs

COTS

AutoSys
Event
Processor

⟷

**AutoSys
Database**

AutoSys is the central
engine of the Processing
CSCI.

Initiates →

Initiates →

Data Management

Data
Manager

Resource
Mgr | Data
Map | DPR

⟷

**PDPS
Database**

**Manages data on Science
Processing HW resources**

**Provides interface to Science
Data Server to acquire (stage)
data or insert (destage) data**

**Manages the retention of
output or input data used by
multiple PGEs**

PGE Execution Management

PGE
Execution Mgr

Resource
Mgr | PGE | PCF

⟷

**PDPS
Database**

Provides PGE exec. prep. activities

Creates Process Control File

**Provides interface to Data Server for
staging of PGE, if necessary.**

**Assists in preparation of science
resources**

Resource Management

Resource
Manager

Resource
Configuration | | Resource

Computer | Disk

⟷

**PDPS
Database**

**Manages Science Processing HW
to support PGE execution**

**Allocates/
Deallocates
Resources** →

**Allocates/
Deallocates
Resources**

Data Pre-Processing CSC

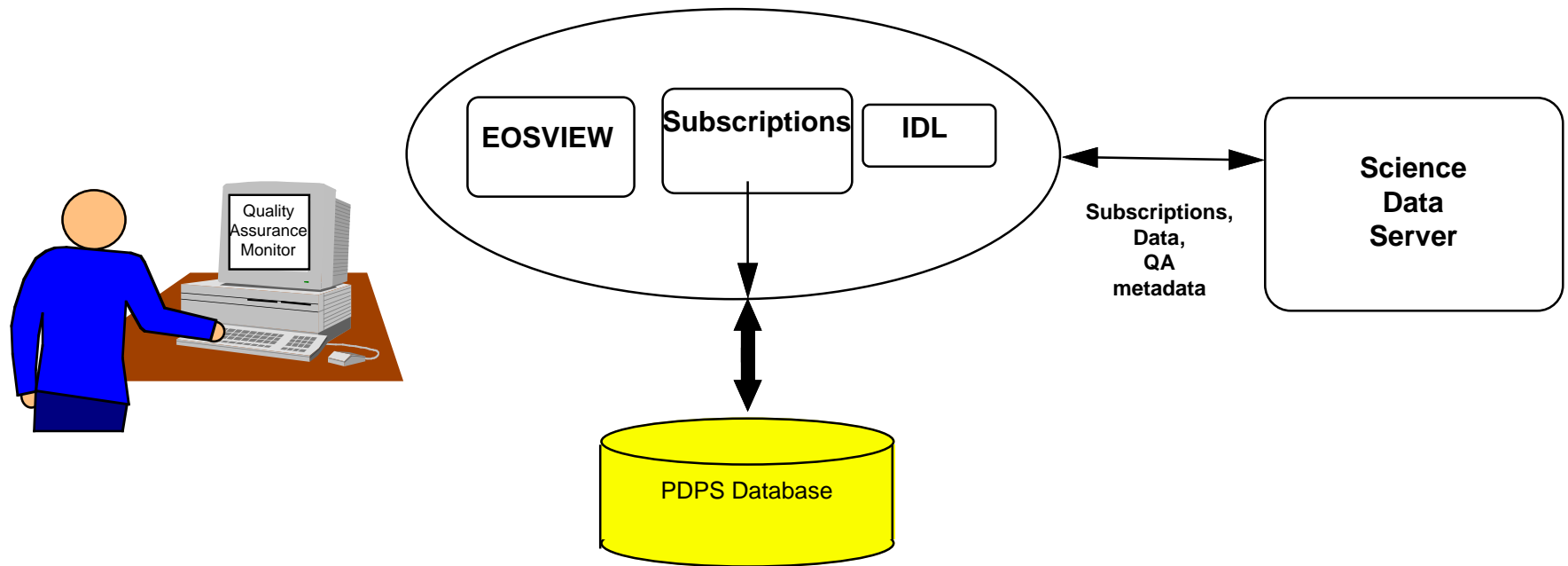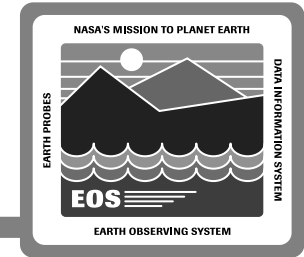At PDR , the Data Pre-Processing functions were grouped together and presented as a separate CSCI.

For CDR, the Data Pre-Processing CSCI is subsumed into the Ingest and Processing CSCI designs.

The Processing CSCI Data Pre-Processing CSC consists of the following:

- Services to reformat Ephemeris data for use by Science Software.
- The Planning CSCI will determine when data pre-processing is required and respond by scheduling the execution of a Pre-Processing PGE.
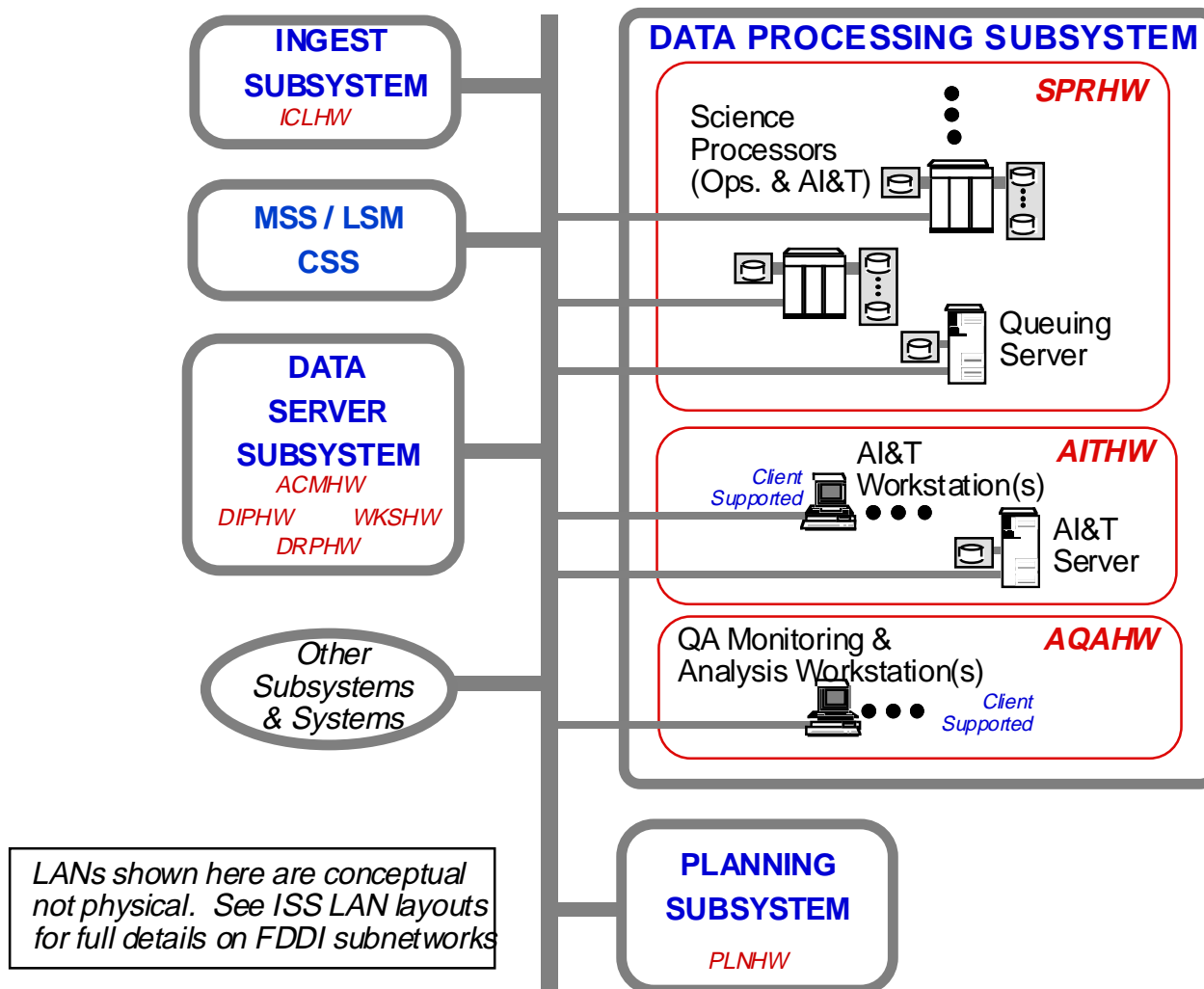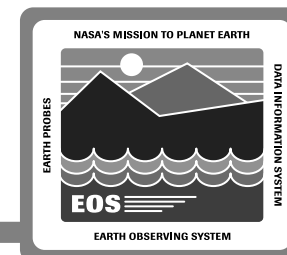
Quality Assurance Monitor CSC

EOSVIEW

Subscriptions

IDL

Quality
Assurance
Monitor

Subscriptions,
Data,
QA
metadata

Science
Data
Server

PDPS Database

Provides Operations Interface to support DAAC QA Activities

- **Data Server Subscription services**
- **Data Staging**
- **Data Visualization**
- **QA Metadata Update**

Data Processing Subsystem Hardware Overview

INGEST SUBSYSTEM
ICLHW

MSS / LSM CSS

DATA SERVER SUBSYSTEM
ACMHW
DIPHW   *WKSHW*
DRPHW

Other Subsystems & Systems

DATA PROCESSING SUBSYSTEM

SPRHW

Science Processors (Ops. & AI&T)

Queuing Server

AITHW

AI&T Workstation(s)
Client Supported

AI&T Server

AQAHW

QA Monitoring & Analysis Workstation(s)
Client Supported

PLANNING SUBSYSTEM
PLNHW

LANs shown here are conceptual not physical. See ISS LAN layouts for full details on FDDI subnetworks

706-CD-001-001 Day 3

305-CD-011-001 Sections 8,9 & 10

SW to HW Mapping

Science Processing
HW Resource

AutoSys Agent
MSS Agent
Data Pre-Processing
Science Software

Science Processing
HW Resource

Science Processing
HW Resource

Science Processing HWCI

305-CD-011-001 Sections 4 & 8

PGE Execution Model

Currently, PGE execution is assumed to occur on a single CPU.
No Timesharing for a CPU is performed.

WHY ?
1) Timesharing provides better CPU utilization for jobs that perform
   significant I/O, but this lowers CPU utilization for CPU-bound jobs.
2) ECS through Release B is not I/O bound in terms of PGE execution.
3) Current AHWGP data used to reach this conclusion.
4) Of the 120 PGEs currently identified through Release B, only 4 are I/O
   bound.

CONSEQUENCES
 This is not a design limitation.

 The allocation of resources is captured at Algorithm Integration and Test
  and made available to Planning. If allocations need to change, this
  information is fed into AutoSys, which will allocate resources according to
  the job definition.

Algorithm Quality Assurance HW
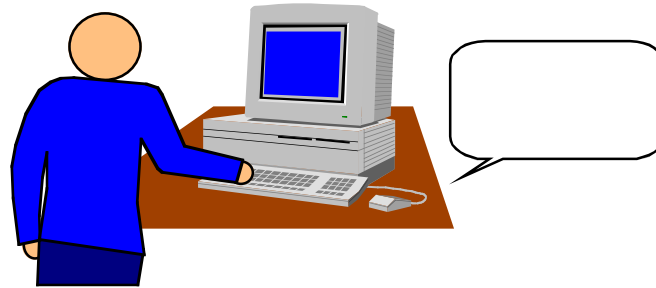
Algorithm Quality Assurance HWCI

Key Assumptions...
- Any Science QA occurring at the DAAC will use similar SW used at the SCF.
- Currently assuming a "One in Five" pull (e.g. 20% checked) to support Science and Non-Science QA.
- Required only at operational sites for Release A (others: AI&T)
- QA process requires some product visualization and/or browse capabilities to check for data integrity.

AQAHW Sizing For MSFC and LaRC
- 20% of Processing Data Inserted into the Data Server is Pulled to the AQAHW
- AQAHW Workstations configured to perform visualization, statistical reporting and trend data storage
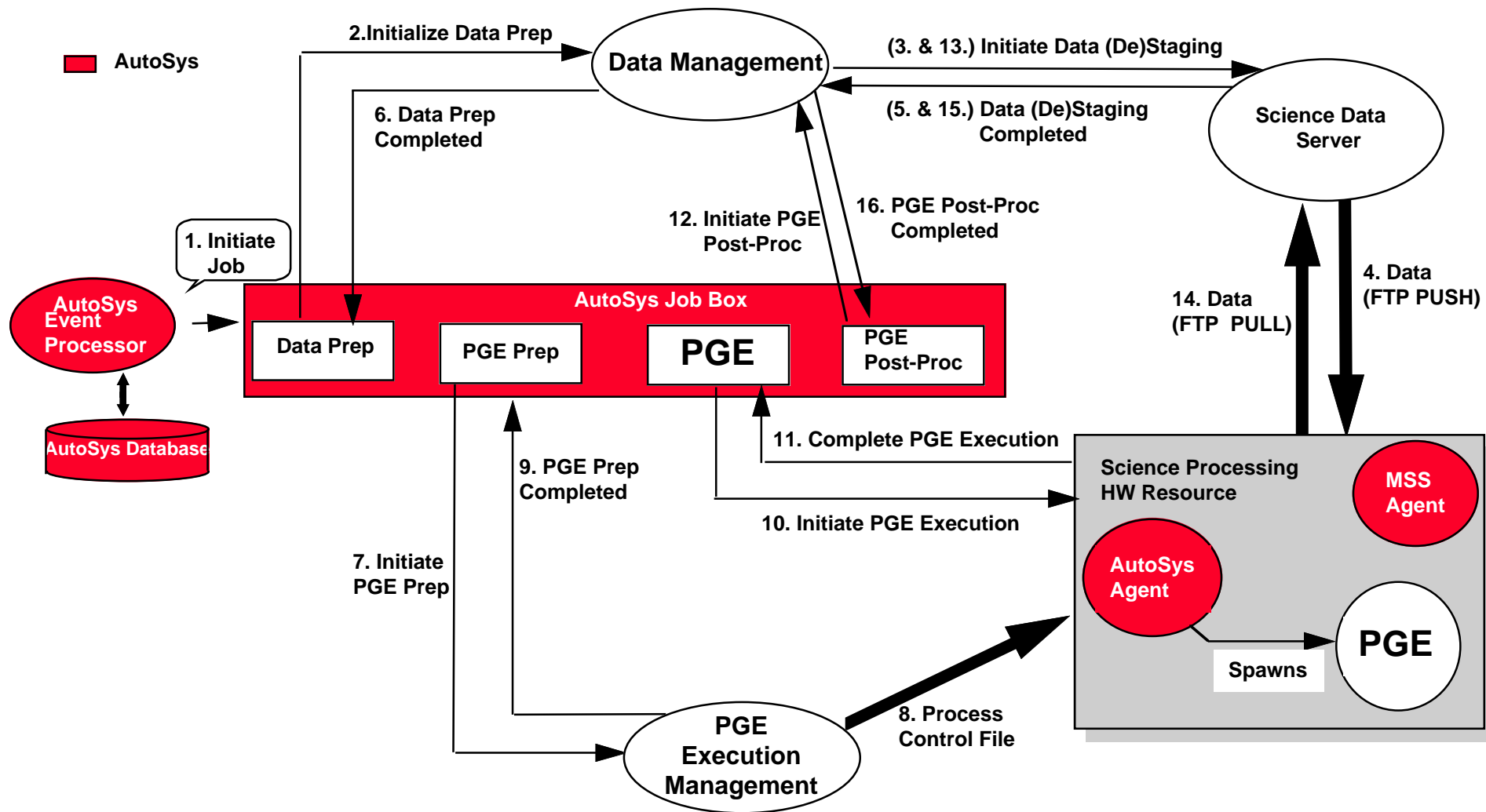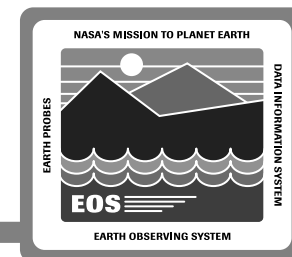- QA Software loads to be refined as QA requirements become clear.
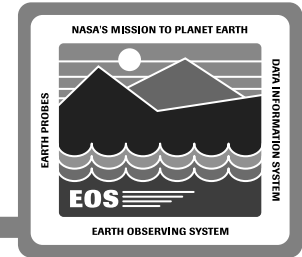
Production Scenario

- **Shows control and data flow paths required to execute a PGE.**

- **Specifies Processing CSCI components role in supporting PGE execution.**

- **Highlights the role of AutoSys and the integration of Processing CSCI and AutoSys components.**

Production Scenario

Production Scenario Exception Handling

Each step of the Production Scenario provides a decision point in responding to errors in the normal production flow.

Each Job will use Status Codes to inform the AutoSys Event Processor of the success or failure of its processing.

The AutoSys Event Processor will determine the recovery steps necessary to respond to the error condition :

- Alert Operations

- Restart Job

- Alert MSS

- Log Error Information

- Destage Failed PGE Information

For Release A, limited automated recovery capabilities will be provided.

DAAC Operations will be responsible in determining what recovery action is necessary.

Evolvability
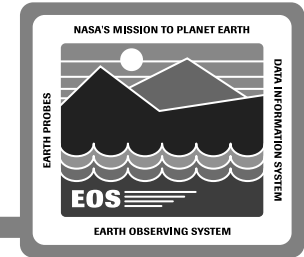
Release B Transition

- AutoSys currently supports Release B processing loads.

- Current customers process 15000 jobs daily in a heterogeneous, UNIX environment.

- Next release will support the ability to share information across DAACs which will enhance ECS cross-DAAC production coordination capabilities.

Parallel Processing

- Prototyping - Science Software Execution Prototype

  Used science algorithms (e.g., Pathfinder AVHRR/Land) to study applicability of various processing alternatives (e.g., SMP, DMP).

- The selection of AutoSys does not hinder ECS' ability to support the parallel processing of PGEs.

Processing CSCI Issues and Recommendations

ISSUE :

DAAC Quality Assurance still not well-defined.

PLAN :

Generic capabilities are provided to perform non-science QA activities.

- **Data Integrity and Checking**
  - Data Visualization Tools
  - Provide mechanisms to check successful insertion of Data Products to the Science Data Server.
  - Provide mechanisms to update associated QA metadata.

If more specific capabilities are needed by the DAACs, this information must be supplied by November 1, 1995.